International Journal of Computer Science Engineering and Information Technology Research (IJCSEITR)
ISSN (P): 2249-6831; ISSN (E): 2249-7943
Vol. 8, Issue 2, Jun 2018, 77-82

TRANS STELLAR
•Journal Publications •Research Consultancy

ENHANCED K STRANGE POINTS CLUSTERING USING BAT INSPIRED ALGORITHM

TERENCE JOHNSON[1], GAURAV CHIBDE[2], POOJA FADTE[3], AMBALI FALARI[4], SAISH GHOLKAR[5] & SUJAY NABAR[6]

[1]*Associate Professor, Department of MCA, Hope Foundation's Finolex Academy of Management and Technology, Ratnagiri, Maharashtra, India*

[2,3,4,5,6]*Students, Agnel Institute of Technology and Design, Assagao, Goa, India*

ABSTRACT

One of the major techniques for data analysis is Clustering in data mining . In this paper, a partitioning clustering method called the Enhanced K Strange Points Clustering algorithm (EKSPA) is used with Bat algorithm. The Enhanced K Strange points clustering algorithm works by first selecting a point that is the minimum (first strange point) of the dataset. It next selects a point that is furthermost (second strange point) from the minimum and continues till it finds K (as many as the number of clusters) strange points which are farthest and equally spaced from each other. The EKSPA then allots remaining points into clusters closest to these K strange points. Finally, it uses the bat algorithm to select the best (bat) point which may replace the K Strange points as the global best solution (bat) if certain conditions are satisfied or retain the K Strange points as the global best solutions (bats) around which the closest points can cluster. As it has been proven that the Enhanced K Strange points clustering algorithm is computationally faster than the K means clustering algorithm while maintaining the quality of clustering, it is concluded that its combination with the bat algorithm also yields better results than the K Means Bat Algorithm.

KEYWORDS: Clustering, Bat Algorithm, Enhanced K Strange Points Clustering, K Means Clustering & K Means Bat Algorithm

Received: Apr 20, 2018; **Accepted:** May 11, 2018**; Published:** May 22, 2018; **Paper Id.:** IJCSEITRJUN20189

Original Article

INTRODUCTION

Data mining techniques are commonly used in the area of financial data analysis, retail industry, telecommunication, industry, biological data analysis, intrusion detection system and other scientific applications. Data mining in simple words can be described as the extraction or exploration of hidden predictive information from massive databases. It is an effective new technology with great ability to help business organizations on being aware about the most essential statistics of their data repositories and warehouses [2]. This information helps the organization to analyse the data and that data can be used for other useful needs. One of the biggest challenges in Data Mining is to choose the right data mining technique. Data Mining technique needs to be selected primarily based on the kind of enterprise and the type of issues faced by the enterprise. A generalized approach has to be used to enhance the accuracy and cost-effectiveness of using data mining strategies. There are basically many techniques and one of the popular techniques used is clustering [1]. Clustering is one of the records analysis strategies which are extensively used in data mining. In this technique, we partitioned the data into a different subset which is known as the cluster. Its main task is exploratory data mining, and a common technique for statistical data analysis used in many fields, including machine learning, pattern recognition, image analysis,

information retrieval, and Bioinformatics [2]. Cluster analysis is one of the prime techniques of data analysis and the K Means clustering algorithm is one that is suitable for grouping a large data sets (MacQueen 1967) [3]. It assigns the number of clusters k and capriciously chooses the initial centroid of each cluster. In order to overcome the problem of improper selection of cluster center in the traditional K-means algorithm which leads to the clustering result into local optimum, the initial clustering center of the K-means algorithm is searched by the bat algorithm. In this paper, the Enhanced K Strange Points Clustering algorithm is used with Bat algorithm for selection of cluster centers around which the nearest points can group. The Enhanced K strange points clustering algorithm computes the K (as many as the number of clusters) Strange points and groups the remaining points in the dataset closest to the computed K Strange points [4]. It then uses the bat algorithm to select the best (bat) point which may replace the K Strange points as the global best solution (bat) if certain conditions are satisfied or retains the K Strange points as the global best solutions (bats) around which the closest points can cluster [4]. In this paper, alternate selection method to select the prospective best cluster centers is done using the bat algorithm. The bat-inspired algorithm, a swarm-based intelligent system impersonates the echolocation system of micro-bats. In the bat-inspired algorithm, the bats randomly fly around the best bat locations found during the search so as to improve their hunting of prey. In practice, one bat location from a set of best bat locations is selected. Thereafter, that best bat location is used by local search with a random walk strategy to inform other bats about the prey location. This paper uses the global-best bat algorithm to select the best bat location [5].

Enhanced K Strange Points Clustering using Bat Inspired Algorithm

The proposed concept of the Enhanced K Strange Points Bat Algorithm (EKSPBA) gives better efficiency and performance compared to the K Means Bat Algorithm (KMBA) [4]. The input to the Algorithm is the iris dataset. As we begin with the bat algorithm, the data points of the iris input set are taken as the bat locations and the other parameters of bat such as frequency, velocity, pulse rate are taken randomly and Enhanced K Strange Points Algorithm (EKSPA) is used to cluster them. As the iris dataset is one which has three flower types, we take K (the number of clusters) equal to 3 [2]. The EKSPA finds three Strange points which are equally farthest from each other by finding the minimum point min, a point at the greatest distance from minimum called max and the third point from a dataset which is equally farthest from min and max. If the third point chosen is closer to either min or max then using equations 1 and 2 we try to bring the third point approximately to the center. These three points, then become the cluster centroids for the three clusters respectively. Further, the closeness of every data point from the dataset is then calculated with respect to strange points and accordingly get assigned to these clusters [4]. Then the bat algorithm is yet again used to update the value according to the best bat solution and gets updated in the dataset. The training dataset consists of 150 records which is being clustered into respective clusters according to the distance between the strange points and gets updated according to bat algorithm. The use of EKSPA in the EKSPBA makes it more efficient than the KMBA [2]. In the KMBA as the dimensions and size of the dataset increases the K Means clustering techniques is likely to take more time to converge as the computation of the next means may take it into an infinite loop. Even if the process is abruptly terminated using ‘t’ the number of iterations in the K Means method, this will lead to inaccurate clusters. These shortcomings can be overcome using the EKSPA [4].

Algorithm

Input: Pre-processed Iris dataset with n objects D={D1,D2,…Dn}

Output: Set of K=3 clusters

1. Initialize the objective function f(x),x = (x1,…xd)T

2. Assign the bat population Xi, where i = 1,2, ... n and Vi

3. Set maximum number of iterations is iter_max

4. Consider t = 1

5. Define pulse frequency fi at Xi

6. Initialize bat position rand, pulse rates ri and loudness Ai

7. While (t <iter_max)

8. Generate a new best solutions by modifiable frequency, for i=l to n do

9. Updating velocities and locations or solutions

Find K_{min}, the minimum of the dataset

Find a point K_{max} which is at a maximum distance from k_{min}

Locate a third point which is farthest from K_{min} and K_{max}

If($d(K_{min},s)==d(K_{max,}s)$)

Kstr = S

else if($d(K_{min},s) < d(K_{max,}s)$)

$$K_{str} = K_{str\ pre} + X_m [\ |K_{max} - Kstr_{prv}\ | / (K-1)\ ] \quad (1)$$

else if($d(K_{max},s) < d(K_{min,}s)$)

$$K_{str} = K_{min} + X_m [\ |K_{str\ prv} - K_{min}\ | / (K-1)\ ] \quad (2)$$

where

K = number of clusters

X_m ranges from 1, 2, 3,…. K-2 i.e

$X_m = X_l, X_2, X_3 \ldots\ldots X_{k-2}$

For e.g. when K=4, $X_m = 4-2 = 2$ so we have

$X_1 = 1$ = for first corrected value of S

$X_2 = 2$ = for second corrected value of S

$Kstr_{prv}$ = uncorrected value of S and

Kstr = corrected value of S

- Repeat the above procedure until we locate K strange points

- Assign the remaining points in the dataset into clusters formed by these non collinear K-Strange points
- Output K clusters

10. If (rand >r_i)
11. Select a solution among the best solutions
12. Generate a local solution around the selected best solution
13. End if
14. Fly randomly and generate new solution
15. if(rand <A_i& $f(x_i) < f(x^*)$)
16. Accept the new solutions
17. Increase the rates and reduce the loudness
18. End if
19. Rank the bats and find the current best x*
20. Increment t
21. End while
22. Post process results and visualization
23. Get the solution from the above bat algorithm and fix the Optimal Location
24. Assign bats (bat locations) closest to these K optimal locations into K clusters
25. Stop

Experimental Result

On executing the EKSPBA it is observed that the cluster centers are obtained faster the KMBA and requires no iterations to calculate the cluster centers as it the case with KMBA. As far as the quality of clustering is concerned, it is observed that the clusters obtained are very close to those obtained with KMBA.

- **KMBA**

```
total number of point
cluster 1 = 38
cluster 2 = 62
cluster 3 = 51
```

- **EKSPBA**

```
position of kmin point      50.5 26.3 15.3 4.3
position of kmax point      85.7 42.8 74.7 25.2
position of kstrange point 58.2 30.15 36.2 12.55
```

```
no of points in cluster 1 : 50
no of points in cluster 2 : 39
no of points in cluster 3 : 61


After clustering (rand(0.2)>r)
bats satisfying condition are :

cluster 1 : Bat42
current best solution in cluster1 : Bat42

cluster 2 : Bat118
current best solution in cluster2 : Bat118

cluster 3 : Bat150
current best solution in cluster3 : Bat150

 Generate new solution by flying randomly

after checking (rand(0.2)<A)&&F(x)<f(x*)=0.8


Accepted solution for cluster1 = Bat42
Accepted solution for cluster2 = Bat118
Accepted solution for cluster3 = Bat150

after decreasing loudness and increasing pulse rate


For cluster1 best bat after ranking is Bat42
For cluster2 best bat after ranking is Bat118
For cluster3 best bat after ranking is Bat150


updated centroids are :
cluster1 : 50.5 26.3 15.3 4.3
cluster2 : 85.7 42.8 74.7 25.2
cluster3 : 58.2 30.15 36.2 12.55
total number of point
cluster 1 = 51
cluster 2 = 39
cluster 3 = 61
```

CONCLUSIONS

A hybrid approach can find more accurate results of clusters. The EKSPBA proposed in this paper yields more accurate and efficient results as compared to the KMBA. Use of EKSPA takes less computation time than the orthodox K Means clustering techniques. One of the limitations is that due to the merging of EKSPA with Bat Algorithm, the computations are much slower than individual basic EKSPA and K Means algorithm. Further work can be done to remove the limitations.

REFERENCES

1. *Xin-She Yang and Xingshi He, "Bat Algorithm: Literature Review and Applications", Int. J.Bio-Inspired Computation, ISSN: 2248-9622, Vol. 5, No. 3, May 2013, pp.141–149.*
2. *Pang-Ning Tan and Michael Steinbach and Vipin Kumar, "K-Means", Chapter 8, Cluster Analysis: Basic Concepts and Algorithms, Introduction To Data Mining, pp.496-508.*
3. *G. Komarasamy, A. Wahi, "An optimized k-means clustering technique using bat algorithm" European Journal of Scientific Research, Vol. 84, No. 2, 2012*
4. *Terence Johnson and Santosh Kumar Singh (2015): "Enhanced K Strange Points Clustering Algorithm", Proceedings of the '2nd International Research Conference on Emerging Information Technology and Engineering Solutions' EITES 2015, 978-1-4799-1838-6/15, IEEE Computer Society Washington, DC, USA © 2015 IEEE, DOI 10.1109/EITES.2015.14, indexed in ACM Digital Library, pp 32-37.*
5. *Bat-inspired algorithms with natural selection mechanisms for global optimization*
6. *Verma, Chandra Prakash, and Neetu Sharma. "Heterogeneous Enhanced Leach Protocol in Wireless Sensor Network to Prolong Network Lifetime with Static Clustering."*
7. *Mohammed AzmiAl-Betar, Mohammed A. Awadallah, Hossam Faris, Xin-She Yang, Ahamad Tajudin Khader, Osama Ahmad Alomari.*